Construction of Unsupervised Prose Text Emotional Lexicon Based on Multidimensional Fusion

Kai Zhang[1,2(✉)], Jianshe Zhou[1,2], and Su Dong[1,2]

[1] Capital Normal University, 105 West Third Ring Road North, Haidian District, Beijing, China
irs_zhangkai@163.com

[2] Research Center For Language Intelligence of China, 105 West Third Ring Road North, Haidian District, Beijing, China

Abstract. Affective computing is an important tool for language processing and opinion mining, and emotional lexicon is the basis of emotional computing, and prose accounts for a large proportion in Chinese teaching and application in China. The construction of special emotional lexicon for prose language learning and language understanding is of great significance to the development of machine assisted human language learning and the improvement of machine deep reading comprehension. Therefore, the research on the construction of prose emotional lexicon is of great significance and value. In this paper, with the help of data collection tools, more than 27000 pieces of modern famous prose database are constructed. After preprocessing the data, denoising, deleting and selecting are completed to determine the walk set. Compared with PMI and word2vec, the accuracy of the method is improved by 16% and 14.8%, which proves that the comprehensive vector space can effectively improve the emotional vocabulary recognition of prose. Finally, 12762 prose general emotional lexicon is formed with the help of this method.

Keywords: Prose emotional lexicon · Prose reading comprehension · Random walk · Word vector · Word co-occurrence

1 Introduction

There has been no definite conclusion on the definition of prose. So far, it is difficult for prose to turn "category" into "body", which is even more difficult in Chinese education. In China's basic education stage, prose learning accounts for a prominent proportion. The number of prose in junior high school Chinese textbooks accounts for "half of the country", which is the most important part of textbook selection. As a literary style, it has its own unique teaching value, which has an indelible role in middle school students' Chinese learning [1]. Senior high school Chinese textbooks are composed of novels, dramas, poems and essays. Taking PEP senior high school Chinese compulsory textbook as an example, according to statistics, there are 65 texts and 11 essays, accounting for 17% [2].

Y. Weng et al. (Eds.): TridentCom 2020, LNICST 380, pp. 134–145, 2021.
https://doi.org/10.1007/978-3-030-77428-8_11

Prose is a literary genre with ideological, narrative and aesthetic features [3]. Compared with other literary styles, prose is more complex, and its emotional expression is particularly profound and changeable, which is a great challenge to the study of prose text understanding. Emotional words are one of the representative research contents in the field of artificial intelligence [4]. Emotional words in prose contain the author's rich emotional information, which plays an important role in prose text understanding. A good emotional lexicon is an indispensable foundation for emotional analysis [5]. The construction of an emotional dictionary in the field of prose can help computers better identify and understand the expression of emotions in prose, and improve the efficiency and accuracy of prose machine understanding.

2 Literature Review

Emotional dictionaries can be divided into general affective dictionaries and domain affective dictionaries. At present, most general affective dictionaries are constructed manually [6]. The more famous English emotion dictionaries are Sentiwordnet, General Inquire and Opinion Lexicon, while the Chinese ones are HowNet, DUTIR, NTUSD, etc. The construction of a general emotional dictionary manually annotates words by mining the synonymous, antonymy and hyponymy relations between words. The construction of general emotional dictionary is more dependent on the integrity of semantic knowledge base [7].

At present, there are relatively few standard Chinese emotion dictionaries. There are three open emotion dictionaries in Chinese information processing. HowNet emotional lexicon is the earliest and most widely spread. There are 4569 and 4370 commendatory and derogatory words in Chinese. The emotional lexicon of Dalian University of technology is divided into 7 categories and 21 sub categories according to the types of parts of speech, emotional categories, emotional intensity and polarity. The emotional intensity is divided into five grades: 1, 3, 5, 7 and 9. 9 indicates the strongest intensity, with a total of 27466 emotional words. There are 11171 Chinese affective dictionaries constructed by Taiwan University. The existing general Chinese emotional lexicon can not accurately judge the author's emotional inclination in the prose environment, and the artificial construction of emotional lexicon consumes a lot of human and material resources, so it is particularly important to automatically build a specific emotional lexicon around Chinese prose. At present, there are three kinds of methods for automatic construction of emotional lexicon: Based on semantic lexicon, based on corpus [8] and the combination of the two [4].

By mining the relationship between words through semantic dictionary database, we can construct emotion dictionary. Rao D et al. [9] extracted positive and negative affective words from WordNet by means of semi-supervised learning method and by giving positive and negative seed sets and synonym graphs. Hu M et al. [10] take adjectives as the research object, construct a seed set of positive and negative emotional words manually, and expand the emotional lexicon through the synonymy and antonymy of words in WordNet, and finally form a large-scale emotional dictionary. Strapparava C et al. [11] added some parts of speech such as nouns, verbs and adverbs on the basis of Hu M, and proposed a comprehensive emotion dictionary based on WordNet under the influence

of multi part of speech. Choi Y [12] proposed a sentiment word construction algorithm based on the perception layer of FrameNet. Jia L [13] used Russell's (1980) model to capture the emotional vocabulary of online social media by deleting low-frequency and non emotional words and retaining repeated words.

Conjunctions and co-occurrence are commonly used in the construction of emotional lexicon based on corpus [6]. The method of conjunctions uses different meanings of conjunctions to construct emotional lexicon. The specific ideas are as follows: the emergence of inflectional words usually represents the change of emotional polarity, which will produce different emotional words, while the juxtaposed conjunctions represent the progressive connection of emotions and usually produce words with the same or similar emotions. Word co-occurrence method means that if two words often appear at the same time, it indicates that the emotional polarity of the two words is similar, so as to judge the emotional similarity between words.

However, in the field of discourse, such as prose, the use of words is flexible, and it is not enough to judge the emotional polarity of words only by conjunctions. However, the deficiency of Chinese semantic knowledge and the limitation of domain make the method based on semantic lexicon perform poorly in constructing domain oriented emotion dictionary. The combination of corpus and semantic knowledge can improve the accuracy of emotion tagging. The typical method is the semi-supervised relation graph method [14,15], which uses the semantic relations in the existing general emotional dictionaries to construct the relationship graph between words, and then uses a graph propagation algorithm to iteratively deduce the emotional tendency of the unknown polarity sentiment words in the corpus, so as to construct a relatively perfect domain emotion dictionary. Bing W et al. [16] proposed an unsupervised emotion classification method based on multi-level fuzzy calculation and multi criteria fusion. The self supervised learning fuzzy classification using labeled training data achieves good experimental results.

3 Design of Construction Method of Prose Emotion Dictionary

This paper proposes an unsupervised construction method of prose emotional lexicon based on multidimensional fusion. Through the data collection of prose website, the basic corpus is collected, and the emotional lexicon of Dalian University of Technology (7 categories, 21 sub categories, 27466 emotion words) is set as the basic emotional lexicon. After de-noising and processing the corpus, multi strategies are integrated to form a seed set. By constructing point mutual information (PMI), word contribution of word vector (word2vec) and joint semantic vector of semantic information of prose text, a comprehensive vector space based on prose corpus is formed. Finally, the random walk strategy is used to obtain the candidate sentiment word set which is similar to or similar to the seed set. And the sentiment lexicon in prose field is completed by using classification algorithm. According to the implementation method of this paper, the main construction process is as follows, as shown in Fig. 1.

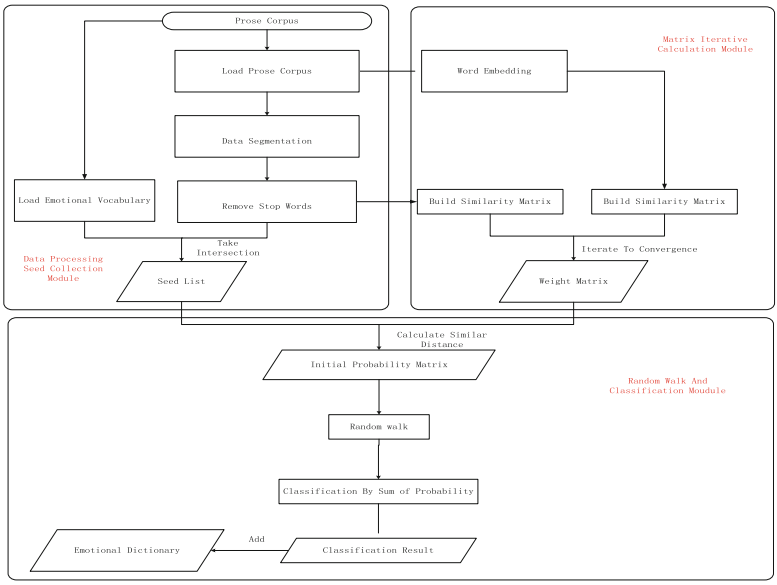

Fig. 1. Flow chart of constructing emotional lexicon of prose text.

The construction process is as follows:

(1) Data preprocessing. It mainly includes the acquisition and selection of prose corpus and the integration of existing emotional dictionaries to form a seed set.
(2) Word vector construction. The word2vec model in deep learning is used to transform words into word vector space, which lays the foundation for subsequent classifiers.
(3) Synthesis vector space classifier design. Through the prose corpus prepared by the author and the emotion dictionary after fusion, the training corpus is constructed. At the same time, the word vector model is used to transform words into comprehensive vector space, and network training is used to generate emotion classifier.
(4) The construction of emotional lexicon is commonly used in prose. The emotion dictionary in the field of prose is mainly composed of two parts: fusion emotional lexicon and candidate emotional lexicon. The classifier model is used to classify and judge its emotional polarity, and finally the common emotional lexicon in the field of prose is obtained.

3.1 Corpus Preprocessing

This study focuses on prose and needs a large number of prose corpus. This paper designs to use the network collection tools to obtain the current network common prose corpus as the candidate corpus. The specific design and collection of prose corpus are shown in Table 1.

It can remove duplicate data and filter ancient poetry. Finally, the collected results are merged into a semi-structured XML document of prose text corpus. The structure contains topics, authors and text. The database collects 27674 masterpieces. The experiment

Table 1. List of collection seed sets of prose corpus.

No.	Name	Link
1	Modern Prose Net	https://www.sanw.net/
2	China Prose Website	https://www.cnprose.com/
3	China Beautiful Article Net	https://www.sengzan.com/
4	Beautiful Article Net	https://www.748219.com/
5	Text Abstract Web	https://www.szwj72.cn/
6	Everyday Beautiful Article Net	https://www.365essay.com/
7	Classic Beautiful Article Website	https://meiwen.ishuo.cn/
8	Short Literature	https://www.duanwenxue.com/
9	Prose Net	https://www.sanwen.net/
10	99 Article Website	https://www.jj59.com/yuanchuang/
11	Prose Online	https://sanwenzx.com/index.html
12	Love net	https://www.biib.cn/
13	Easy Prose Reading Website	https://www.cnease.cn/index-htm-m-bbs.html

uses Python's beautiful soup library to complete the analysis, uses Jieba word segmentation and completes the stop word operation through the Harbin Institute of technology's stop list, and obtains 9449438 prose corpus. A total of 7012 affective words were selected from the intersection of the thesaurus and the Chinese emotional vocabulary ontology database (Dalian University of Technology). After classifying the emotional lexicon into seven categories: sadness, evil, good, surprise, fear, anger, and happyness, 10 emotion words were randomly mapped and selected as seed sets.

3.2 Lexical Relevance Analysis of Prose Corpus

By analyzing the content of prose corpus, it is easy to get that the words in prose are more implicit and profound in expressing emotions. All scenery words are emotional words, and different scenes express different emotions. Generally, not only adjectives express the author's emotion, but also the collocations of adverbs and adjectives contain emotional tendency. The connection of emotional words in prose and the rules that prose writers want to express their emotions are largely implied in the corpus. Therefore, if we want to obtain the emotional words, we need to consider the correlation between prose words.

PMI (pointwise mutual information) is a special case in NMI (normal mutual information), and NMI is a concept derived from information theory, which is mainly used to measure the degree of correlation between two signals. PMI is used to calculate the degree of correlation between two words in text processing. Compared with traditional similarity calculation, the advantage of PMI is to find out whether there is semantic correlation or topic correlation between words by finding the cooccurrence of words from the statistical perspective.

Definition 1: Co-occurrence calculation (G_PMI) uses the mutual information between any candidate words to measure the correlation between words. Assuming that the candidate lexicon is C, w_i and $w_j(i, j \subset C)$ is calculated as follows:

$$G\_pmi = \log(\frac{p(w_i, w_j)}{p(w_i) * p(w_j)}) = \log(\frac{p(w_i|w_j)}{p(w_i)}) = \log(\frac{p(w_j|w_i)}{p(w_j)}) \tag{1}$$

If w_i and w_j are distributed independently, then $p(w_i, w_j) = p(w_i) * p(w_j)$, which means $pmi(w_i, w_j) = \log 1 = 0$. On the contrary, if the distribution between w_i and w_j are not independent, then $p(w_i, w_j) > p(w_i) * p(w_j)$, and as the relevance is increased, the value of pmi is greater, which lead us to the conclusion that the more information w_i and w_j carry together, the more easier they will appear together.

Learning semantic knowledge by an unsupervised way in a large number of prose corpora is of great significance to the construction of prose emotional lexicon. At present, the commonly used method is to use vectors to represent semantic relations, Word2Vec uses word vectors to represent the semantic information of words by learning the text, in other words, through an embedding space to make words, which are semantically similar, close in the space. Therefore, this article uses the Word2Vec model to construct the prose vocabulary semantic network.

Definition 2: The semantic word vector space C_w2v is used to represent the contextual semantic relationship between words. The method is to embed a space, which is low-dimensional, to make semantically similar words very close in the space.

$$C\_w2v = \sum_{k=1}^{n} S(x_i, x_j) = \frac{\sum_{k=1}^{n} (x_{ik} x_{jk})}{\sqrt{\sum_{k=1}^{n} x_{ik}^2} \sqrt{\sum_{k=1}^{n} x_{jk}^2}} \tag{2}$$

$x_i \in R, x_j \in R$, R is the corpus set, and the dimension is n-dimensional, at which the larger the $s(x_i, x_j)$, the greater the similarity between the two words. Skip-gram is an unsupervised learning method that can be used for any original text. Compared with other word-to-vector expressions, it requires less memory. Only two weight matrices with dimensions [N, |v|] instead of [|v|, |v|] are needed. However, the calculation of the Softmax function of this model is time- consuming.

Definition 3: By Definition 1 and Definition 2, the co-occurrence relationship calculation G_pmi and the semantic word vector space C_w2v can be obtained respectively. In order to integrate the co-occurrence calculation and the semantic word vector space, the comprehensive vector space E (G_pmi, C_w2v) is specially defined.

$$\text{Vertical iteration} : E_v(t+1) = \partial \times S \times E_V(t) + (1 - \partial) \times C_{w2v} \tag{3}$$

$$\text{Horizontal iteration} : E_h(t+1) = \partial \times E_h(t) \times S + (1 - \partial) \times E_v^* \tag{4}$$

As mentioned above, the ∂ is the propagation parameter, and $\partial \in (0, 1)$, $E_{\mathrm{v}}(0) = D$, the matrix is $S \subset R^{m \times m}$, $S = Z^{-1/2}\mathrm{G\_pmi}Z^{-1/2}$

Firstly, by using Definition 2 to continuously iterate to get the adjacency column matrix E_{v}^{*}. Then the horizontal iteration is carried out until it tends to be stable, and get the adjacency matrix E_h^{*}. Finally, the integrated vector space E is obtained.

3.3 The Construction of Common Emotional Word Database in Prose

The random walk model [17] describes the correlation between candidate words through the connectivity between points, the model falls into two categories, random walk graph and random walk process. The random walk process is based on the random walk graph, setting out from the word x of unknown emotional tendency, then starting to walk, among all the words connected to the word x, if a word is more closer to the word x on the model graph, then the probability of getting to this word will be greater, and vice versa. Briefly speaking, the connection probability between points is used to measure the random walk correlation between words, which is suitable for the global exact minimum and the opposite of gradient descent.

For the $\forall x$ in the graph, which jumping to any vertex $\forall y$ with probability $\mathrm{p}(x, y)$, the $\mathrm{p}(x, y)$ is called the jumping probability. The walking process requires 4 input parameters, adjacency matrix E_h^{*}, initial probability distribution vector S_0, jump occurrence weight β and random jumping probability p. The random process iteration formula is as follows:

$$S_{\mathrm{i}}(t+1) = (1-\beta) \times E_h^{*} \times S_i(t) + \beta \times p \ (0 < \beta < 1) \tag{5}$$

The vector $S_i(0)$ is the initial probability distribution from the word X_{i} to the seed set, m is the number of seed sets. Through the iteration of formula (3), a stable probability distribution $S_{\mathrm{i}}*$ can be obtained, and then the probability distribution of each type of seed set is calculated and the maximum class probability is obtained, that is, the maximum is taken after the sum of each seed probability of the type of seed set.

$$O(x_i[c]) = \arg\max_{0<c<8} \sum_{j=1}^{n} x_{ij}[c] \tag{6}$$

Among them, $\mathrm{O}(\mathrm{x}_i[c])$ represents the probability from the word x_i to the largest seed category C, and C represents the seed category, which is taken $\mathrm{c} \in (0, 8)$ here, since the emotional word category is divided into seven categories, j represents the seeds in the seed set $\mathrm{j} \in (1,\mathrm{n})$, and the n represents the number of seeds in each category.

4 Experimental Results and Analysis

According to the construction process of prose emotional lexicon as shown in Fig. 1, the server is configured as Intel (R) Xeon (R) Gold 5118 CPU @ 2.30 GHz, memory 256G.

4.1 Experimental Results of Data Preprocessing

After collecting the link organization in Table 1, 27674 famous prose works were obtained after de duplication and filtering. After word segmentation and elimination of stop words, 4718244 were obtained and 77389 were de duplicated. The thesaurus is deleted according to the frequency of words, and 10 times is selected as the threshold. The candidate word sequence is obtained by taking the intersection with the Chinese emotional vocabulary ontology database (Dalian University of technology emotional lexicon), with a total of 7012 emotional words. According to this candidate sequence, the sentiment types and parts of speech are analyzed as follows (Fig. 2 and Table 2):

Table 2. Analysis of emotional types and part of speech distribution of candidate word.

No.	Classification	Numbers	No.	Classification	Numbers
1	Happiness	656	1	Noun	2591
2	Evil	2244	2	Verb	2879
3	Good	3167	3	Adjective	1210
4	Grief	522	4	Adverb	128
5	Fear	275	5	Other	204
6	Surprise	66			
7	Anger	82			

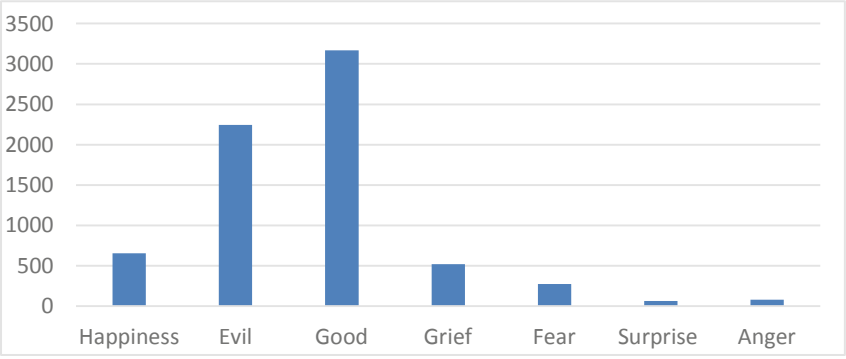

Fig. 2. Distribution of emotional words.

According to the analysis of the above results, the good and the evil are the same in the prose corpus, and there are few articles about anger. From the candidate parts of speech, verbs and names are more distributed. In order to balance the number of different kinds of seed sets, good experimental results are achieved. According to the above experimental data, the number of seed sets was selected. The accuracy rate (P), recall rate (R) and F1 value were used to evaluate the experimental results.

Table 3. Comparison experiment of seed set number selection.

The number of seed words	Accuracy	Recall	F1 value
6	35.3%	38.2%	36.1%
8	49.1%	49.7%	40.4%
10	56.8%	51.6%	55.4%
12	53.2%	50.6%	51.3%
14	51.4%	52.6%	50.7%

According to the experimental results of Wang [4] and others on low-frequency words, we combine the experimental results in Table 3. The analysis shows that the accuracy rate, recall rate and F1 value are the optimal solutions when there are 10 kinds of seed set respectively. Therefore, after mapping the emotional lexicon into seven categories: sadness, evil, good, surprise, fear, anger, and happiness, 10 emotion words are randomly selected as seed set (Table 4).

Table 4. Seed set of each emotion label in the random walk model

No.	Category	Seed set
1	Grief	血癌, 渺茫, 落难, 怏怏, 挫败, 变迁, 倾家荡产, 瘦瘠, 断气, 道别, 遇害
2	Evil	闪失, 不良, 旷日持久, 违反, 失禁, 驱使, 失效, 烦躁, 诬告, 喝斥
3	Good	振奋人心, 飘逸, 财礼, 标致, 憨笑, 亲历, 孝子, 见义勇为, 晶亮, 回报
4	Surprise	突如其来, 不期而然, 怔住, 惊天动地, 大吃一惊, 目瞪口呆, 奇特, 奇怪, 轰动, 天崩地裂
5	Fear	急切, 急如星火, 囚笼, 汗流满面, 危机四伏, 慌乱, 心悸, 交战, 汗颜, 扑腾
6	Anger	赌气, 怒冲冲, 不甘示弱, 惩治, 咆哮, 气急败坏, 泄愤, 可气, 疾言厉色, 针锋相对
7	Happiness	秋波, 水落石出, 朝阳, 尽意, 欢欣鼓舞, 启示, 启发, 畅顺, 畅所欲言, 盈余

4.2 Vocabulary Relevance Experiment of Prose Corpus

In order to achieve the goal of constructing prose emotional lexicon effectively and automatically, three groups of comparative experiments are designed to verify the accuracy and scalability of the method. According to the definition of correlation analysis in 3.2, G_ pmi, C_ w2v, E(G_ pmi, C_ W2v) can be obtained, it is co-occurrence relation matrix, semantic word vector space and comprehensive vector space.

According to the experimental results of the three methods, the methods are evaluated from the recognition accuracy of prose emotional lexicon and the classification accuracy

of emotional words. The evaluation was carried out by random sampling method. Multiple groups of comparisons were set and the final mean value was taken as the global effect.

Classification accuracy of prose emotional lexicon: according to the results of the three methods, 30 emotion types were randomly selected from each category, and 7 groups of 30 words in each group were constructed as evaluation samples. After judging whether the sample belongs to the corresponding emotion category, the probability is obtained. $p(i) = \frac{q}{m}$, i is the type of emotion, q is the correct number of recognition, m is the number of samples.

Table 5. Accuracy evaluation of emotional lexicon classification in prose

Algorithm	Happiness	Anger	Grief	Good	Surprise	Fear	Evil
C_w2v	24.8%	10.7%	20.2%	30.9%	13.7%	30.7%	16.5%
G_pmi	20.6%	10.4%	17.8%	20.6%	17.8%	7.9%	10.3%
E (G_pmi,C_w2v)	**42.7%**	**32.3%**	**37.8%**	**47.5%**	**44.2%**	**39.9%**	**35.8%**

As can be seen from Table 5, the classification accuracy of emotion words in this paper has been significantly improved in seven dimensions. The second is C_W2v method, PMI method is the worst. It can be seen that in the field corpus such as prose, emotional expression is more diversified, rhetorical devices are widely used, and semantic similarity can effectively improve the identification of emotional words.In addition, the accuracy rate of "anger" in emotion category is low, which is due to the close relationship between "sadness" and "surprise". For example, "mourning without dispute" has both anger and sadness, resulting in the low accuracy of recognition.

The accuracy rate of emotional lexicon recognition: from the three groups of experimental results, 100 emotion words were randomly selected according to 5 groups. Through the evaluation of the results, we can judge whether they are emotional words. $p(k) = \frac{q}{m}$, k represents the group, q is the correct number of recognition, and m is the number of samples. Finally, the average value of the accuracy rate is calculated as the evaluation of the overall accuracy effect of this method.

Table 6. Evaluation of the recognition accuracy of prose emotional lexicon.

Algorithm	Group 1	Group 2	Group 3	Group 4	Group 5	Average
C_w2v	30.9%	39.4%	24.3%	32.8%	32.9%	32.0%
G_pmi	40.1%	32.5%	35.2%	29.7%	28.8%	33.2%
E (G_pmi,C_w2v)	47.6%	42.7%	45.1%	53.5%	51.3%	**48.0%**

It can be seen from Table 6 that the average recognition accuracy of PMI's prose emotional lexicon is higher than that of word2vec, which indicates that the co-occurrence

of words in prose genre can better reflect the relevance between words, and it plays a greater role in the identification process of prose emotional words. The comprehensive vector space strategy proposed in this paper can effectively integrate the advantages of the two. With the help of random walk strategy, the local optimal solution is obtained, which is the overall optimal emotion word recognition effect, and the recognition effect is improved by 14.8%.

From the result table of correct emotion words, we can see that the highest number of emotion words identified by this algorithm is 12762, 10433 are identified by PMI, and 6824 are identified by word2vec. To sum up, PMI is better than word2vec in emotional word recognition, but it is not as accurate as word2vec in the classification and recognition of emotional words. From the experimental results, the proportion of "happy" and "good" texts is larger. At the same time, in the prose corpus the emotional color of the text is mostly commendatory, expressing positive emotions, or expressing sad emotions, with less other types.

5 Summary

In this paper, a domain corpus was constructed by proses. Firstly, the cohesion between words is fully considered. Secondly, the seed set is selected randomly by analyzing the category, part of speech and frequency of emotional words. Thirdly, the co-occurrence degree is constructed by PMI and word2vec algorithm respectively G_pmi matrix and C_w2v matrix. Then, the synthesis vector space E is constructed. After data validation, the method has good discrimination effect, effectively constructs the field corpus of prose field, and the experimental results verify the effectiveness of the proposed method.

Prose expresses the author's true feelings, flexible writing style, more diversified forms of expression, more subtle and subtle emotional expression. Due to the differences in personality, life experience and life attitude, the emotional types and expression ways of writers often have their own characteristics. Therefore, it puts forward better requirements for fine classification of emotional lexicon. At the same time, the use of rhetorical devices in prose is high. The machine understanding and processing of these rhetorical devices plays an important role in the automatic construction of emotional lexicon. For example, in Yu Dafu's autumn of the old capital, "I want to have a good taste of the 'autumn' of the old capital".

Therefore, the author of this paper will focus on the rhetoric recognition of prose and the detailed classification of emotional words in proses, so as to further enrich the emotional lexicon of prose and promote the discovery of new words and emotional connotation of prose emotional words.

Acknowledgement. This research was financially supported by "Research on key technologies and model verification of prose genre oriented text understanding (ZDI135–101)", "Research and Application of Key Technologies of Intelligent Auxiliary Reading System (ZDI135–79)", Capacity Building for Sci-Tech Innovation-Fundamental Scientific Research Foundation (20530290082).

References

1. Hui, G.: Junior High SchoolTextbook "Unified Edition" and "Curriculum Standard Edition" The Same Modern Prose Items Comparatively Study. Yan'an University (2020)
2. Lina, M.: Research on the Teaching Strategies of Modern and Contemporary Prose in Senior High Schools under the Background of New Curriculum Standards. Shaanxi Normal University (2019)
3. Zhaoshegn, W., Ping, H., Jiang, L.: The definition of prose. People's Daily Overseas Edition.
4. Suge, W., Qi, C., Xin, C.: Prose-oriented low-frequency emotion word extraction and emotion label determination. J. Shanxi Univ. (Nat. Sci. Ed.) **42**(02), 321–331 (2019)
5. Lili, M., Heyan, H., Xinyu, Z.: Overview of the construction of emotional dictionaries. J. Chin. Inf. Process. **30**(5), 19–27 (2016)
6. Ke, W., Rui, X.: A review of automatic construction methods for emotional dictionaries. Acta Automatica Sin. **42**(04), 495–511 (2016)
7. Fenglin, L., Yaxian, F.: Research on the Construction Method of Domain Emotion Dictionary. Libr. Theor. Pract. **2019**(12), 60-65+112 (2019)
8. Jiaheng, H., Yonghua, C., Chengyao, W.: Automatic construction of domain sentiment dictionary based on deep learning-take the financial domain as an example. Data Anal. Knowl. Disc. **2**(10), 95–102 (2013)
9. Rao, D., Ravichandran, D.: Semi-supervised polarity lexicon induction. In: Proceedings of the 12th Conference of the European Chapter of the Association for Computational Linguistics, DBLP 2009, Athens,Greece, pp. 675–682 (2009)
10. Hu, M., Liu, B.: Mining and summarizing customer reviews. In: Proceedings of the 10th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 168–177. ACM (2004)
11. Strapparava, C., Valitutti, A.: WordNet affect: an affective extension of WordNet. In: Proceedings of the 4th International Conference on Language Resources and Evaluation (2004)
12. Choi, Y., Wiebe, J.: +/−EffectWordNet: sense-level lexicon acquisition for opinion inference. In: Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing, pp. 1181–1191 (2014)
13. Zhao, J.L., Li, M.Z.: The development of the Chinese sentiment lexicon for internet. Front. Psychol. **10**, 2473 (2019). https://doi.org/10.3389/fpsyg.2019.02473
14. Fang, W., Yong, H.: Towards building a high-quality microblog-specific Chinese sentiment lexicon. Decision Support Syst. **87**, 39–49 (2016)
15. Khan, J., Lee, Y.-K.: LeSSA: a unified framework based on lexicons and semi-supervised learning approaches for textual sentiment classification. Appl. Sci. **9**(24), 5562 (2019). https://doi.org/10.3390/app9245562
16. Bing, W., Wei, H.: An unsupervised sentiment classification method based on multi-level fuzzy computing and multi-criteria fusion. IEEE Access **8**, 422–434 (2020)
17. Abdaoui, A., et al.: FEEL: a French expanded emotion lexicon. Lang. Resour. Eval. **51**(3), 1–23 (2017)